DOCUMENT RESUME

ED 167 632 TM 008 469

AUTHOR	Floden, Robert E.; And Others
TITLE	Don't They All Measure the Same Thing? Consequences of Selecting Standardized Tests. Research Series No. 25.
INSTITUTION	Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C. Basic Skills Group. Teaching Div.
PUB DATE	Jul 78
CONTRACT	400-76-0073
NOTE	21p.; Paper presented at the Winter Conference on Measurement and Methodology (Los Angeles, California, 1978) ; Parts marginally legible due to print quality
AVAILABLE FROM	Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, Michigan 48824 ($1.50)
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	*Achievement Tests; *Curriculum; *Elementary School Mathematics; Grade 4; Intermediate Grades; Item Analysis; Standardized Tests; Teacher Response; *Testing Problems; Test Reliability; *Test Selection; *Test Validity
IDENTIFIERS	Comprehensive Tests of Basic Skills; *Content Validity; Iowa Tests of Basic Skills; Metropolitan Achievement Tests; Stanford Achievement Tests

ABSTRACT

The authors argue that personnel who select standardized achievement tests have been led to believe that the major achievement test batteries differ very little in terms of the topics they test; but that the content covered by these major tests is different, and that such differences have consequences for instructional content. To test this argument, four fourth grade mathematics achievement tests were examined: the Stanford Achievement Test (SAT), Iowa Tests of Basic Skills (ITBS), Metropolitan Achievement Tests (MAT), and the Comprehensive Tests of Basic Skills (CTBS). Test content was compared to topics which a teacher might cover. In the "operations" category, the tests corresponded quite closely in the percentages of items involving certain adding, subtraction, and division tasks. However, important differences were found in other categories and the number of items covering addition, grouping, alternative number systems, and percentages varied considerably between tests. Evidence of high internal consistency or test reliability does not rule out the effects of item-treatment-interactions. It is suggested that the problem of teachers teaching only the content covered by the test is not as great as might be thought. The authors recommend that tests be selected consistently and that their content be valid for the curriculum content. (Author/GDC)

ED167632

TM008 469

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Lawrence W. Lezotte

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Institute for Research on Teaching

College of Education — Michigan State University

Research Series No. 25

DON'T THEY ALL MEASURE THE SAME THING? CONSEQUENCES OF SELECTING STANDARDIZED TESTS

Robert E. Floden,
Andrew C. Porter, William H. Schmidt,
and Donald J. Freeman

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

July 1978

The work reported herein is sponsored by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Teaching Division of the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-76-0073)

Don't They All Measure the Same Thing? Consequences of Selecting Standardized Tests[1]

Robert E. Floden,
Andrew C. Porter, William H. Schmidt,
and Donald J. Freeman[2]

Recently, several authors have called attention to the importance of consistency between the content of instruction and the content of tests used to assess instruction (e.g., Walker & Schaffarzick, 1974; Wiley, Note 1; Porter, Schmidt, Floden, & Freeman, Note 2). If test content does not match instructional content, test results might reflect a distorted picture of the instruction effects. Many tests, it has been found, cover only a fraction of the content presented in many elementary classrooms. The major standardized achievement tests for elementary schools, for example, focus on basic skills, and, as a result, the test scores reflect only achievement in basic skills, rather than achievement in the total instructional content covered.

Many people acknowledge that standardized test scores do not reflect instruction in content outside the basic skills, but it has seldom been recognized that even the definition of basic skills may vary from test to test. In fact, the tests do not all measure the same content knowledge, despite the prevailing opinion to the contrary (e.g., Cooley & Lohnes, 1976). This suggests that particular attention should

---

[1]This paper was presented at the UCLA Center for the Study of Education Winter Conference on Measurement and Methodology, 1978.

[2]The authors are senior researchers in the Institute for Research on Teaching and members of IRT's External Factors Research Group.

be paid to the content of the tests chosen to ensure that test results reflect achievement on the content considered important in a local school district.

While the effect of test content on test scores has been recognized, the effect of test content on instruction has received little attention. Although some people believe that the initiation of a testing program may have global effects on instruction (leading perhaps to greater emphasis on the basic skills), more subtle shifts in instructional content brought on by use of particular tests are only beginning to be investigated. These and other influences on a teacher's selection of instructional content deserve careful investigation.

District Test Use

Many school district administrators are concerned with raising student achievement in the basic skills. This concern is reflected by such administrative policies as constructing lists of objectives, meeting with teachers in workshops to discuss goals and methods, rewarding school building administrators for improved performance in their schools, and testing student progress on a regular basis. Although these actions are frequently coupled with the development of criterion-referenced tests, nearly all districts continue to administer norm-referenced standardized achievement tests to assess improvement within the district and to compare their district performance with national norms.[3]

Little is known about the criteria administrators use to select a test series. Factors considered probably include cost, ease of

[3]The terms *criterion-referenced* and *norm-referenced tests* have taken on a variety of meanings. In this context, we define criterion-referenced tests as those in which an individual is assessed relative to a certain standard, whereas norm-referenced tests assess the individual's performance relative to other individuals or to a group average.

administration, ease of scoring, and reporting format. One factor which does not receive enough attention is the specific content covered by each subtest. School personnel have been led to believe that the major test series differ little in terms of the topics they test (at least for subtests with similar titles). High correlations among the tests, together with apparent unidimensionality of each subtest, have suggested that the tests all measure the same things. If that were the case, the content covered by a given test could not be a test-selection criterion, since that content would be almost identical to the content covered by other tests. In this paper, we argue that there are differences in the content covered by the major tests and that those differences have consequences, not only for assessing district progress, but also for the content of instruction in the district's classrooms.

Test Analyses

Four test series dominate the market in elementary reading and mathematics. These series are: the Stanford Tests of Basic Skills (SAT), the Iowa Tests of Basic Skills (Iowa), the Metropolitan Achievement Tests (MAT), and the CTB/McGraw-Hill Comprehensive Tests of Basic Skills (CTBS). Each test series is composed of tests designed for specific levels of achievement, roughly corresponding to grade levels. Tests at each level consist of several subtests, broadly grouped into areas covering different reading and mathematics skills. In mathematics, for example, the subtests typically include "Mathematics Applications," "Computation," and "Concepts."

We at the Institute for Research on Teaching (IRT) recently examined

the fourth-grade mathematics content covered by the four test series.[4] We used an iterative process of analysis and classification of items on the SAT to develop a taxonomy for describing the tests. A complete presentation of the taxonomy is given in Figure 1; each cell in the classification matrix corresponds to a topic that a teacher might elect to cover. (The process of test description has been described elsewhere in greater detail /Porter et al., Note 2; Schmidt, Porter, Floden, & Freeman, Note 3/).

Comparisons of the four tests, detailed in Table 1, indicate that although they are quite similar in some respects, they also have striking differences. On the "operations" dimension, the tests corresponded quite closely in the percentages of items involving "subtract without borrowing " (6-8%), "add or subtract fractions without a common denominator" (0-2%), and "divide with remainder" (1%). For the other levels, however, differences were apparent. Twenty-one percent of the items on the MAT, for example, involved addition, while the corresponding figures for the Iowa, SAT, and CTBS were only 12, 13, and 14% respectively. The Iowa had at least 5 percentage points fewer multiplication items than the other tests. Grouping was tested by the SAT but not by the MAT or the CTBS.

With respect to the nature of the material, more similarities than differences were found, but the differences may be quite significant. Six percent of the items on the CTBS, for example, involved percentage, while the Iowa and MAT had no such items. The MAT and SAT both included items on alternative number systems, while the Iowa and CTBS did not.

---

[4]Copyright dates of the tests which were analyzed are as follows: SAT - 1973; Iowa - 1971; MAT - 1970; CTBS - 1968.

Classification of _____

By _____ Date _____

MODE OF PRESENTATION

Nature of the Material / Operation	Graphics, Figures, Tables or Physical Objects												Operation(s) Specified												Operation(s) Not Specified (Story Problems)											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Whole Numbers — single digits																																				
Whole Numbers — single digit and multiple digit																																				
Whole Numbers — multiple digits																																				
Fractions — single																																				
Fractions — multiple																																				
Decimals																																				
Percents																																				
[a] Alternate Number Systems																																				
Place Value																																				
Number Sentences																																				
[b] Algebraic																																				
[a] Essential Units of Measurement																																				
[a] Geometric Figures																																				
Other																																				

Operations

1. Add
2. Subtract w/o borrowing
3. Subtract with borrowing
4. Add or subtract fractions
5. Multiply
6. Divide w/o remainder
7. Divide with remainder
8. Combination
9. Grouping
10. Identify Equivalents
11. Identify Rule (Order)
12. Identify Terms*

* Be sure to *identify specifics* on attached page.

Figure 1.

Table 1

ITEM DISTRIBUTIONS FOR EACH FACTOR ACROSS TESTS*
FOURTH GRADE LEVEL

	IOWA	MAT	SAT	CTBS
I. Mode of Presentation				
- graphs, figures, tables, etc.	43	15	21	19
- operation(s) specified	29	52	53	59
- operation(s) not specified	29	32	27	22
	(N=84)	(N=115)	(N=116)	(N=113)
II. Nature of Material				
- single digits	12	15	20	2
- single and multiple digits	12	20	23	18
- multiple digits	24	19	22	19
- total -- whole numbers	47	54	66	39
- single fraction	6	4	5	7
- multiple fractions	5	3	-	7
- decimals	6	5	4	10
- percents	-	-	1	6
- alter. number systems	-	2	1	-
- place value	8	3	5	4
- number sentences	6	1	2	-
- algebraic sentences	8	10	8	12
- essen. units meas.	10	15	7	11
- geometric figures	2	3	3	2
- other	1	1	-	2
III. Operations				
- add	12	21	13	14
- subtract w/o borrowing	8	8	6	8
- subtract with borrowing	11	11	6	5
- add or subtract fractions w/o common denominator	1	-	-	2
- multiply	11	19	16	17
- divide w/o remainder	6	9	15	14
- divide with remainder	1	1	1	1
- combination	8	6	7	7
- grouping	2	-	5	-
- identify equivalents	20	18	16	15
- identify rule (order)	11	3	9	12
- identify terms	8	5	6	4

*entries are percents

The significance of these differences becomes more apparent when one considers that a single correct answer on the SAT math subtest can add approximately .2 to a student's grade equivalent score, if it is near the middle of the norm distribution. Since improvements of a fraction of a grade level are generally considered important, it seems likely that the differences among tests could result in score differences that would also be considered important.

Although our analysis of the mathematics tests may be open to criticism, it does illustrate that differences can and sometimes do exist among the standardized tests, tests often thought to be virtually interchangeable. Hence, as we have previously suggested, content covered is a factor which administrators must consider when selecting standardized tests. A school district that emphasizes work with percentages in fourth grade would get a distorted picture of progress from the Iowa, which contains no percentage problems. On the other hand, a district which does not introduce percents until the sixth grade would be unnecessarily discouraged by the results of the CTBS, which contains 6% problems involving percentage.

Differences in supposedly equivalent tests have been found by other authors as well. Linn and Slinde (1977), for example, found that a change in test forms can lead to substantially different grade equivalent socres for low achieving children. Linn (Note 4) also discovered that even when tests have been empirically equated, the choice of test can greatly influence the estimate of program effects. Neither of these studies, however, attempted to explain the cause of discrepancy in results.

Differences in test content may partially explain the discrepancy. The content discrepancy may result, not from the test publishers' failure to include the appropriate content, but from a general lack of agreement on the definition of basic skills. In mathematics, for example, considerable attention was given during the 1960s to the question of whether or not material such as "elementary set theory" was part of the basic skills. That question--and others like it--was never answered, as evidenced by proceedings of a National Institute of Education (NIE) Conference in October 1975, directed at the question, "What are basic mathematical skills and learning?" (NIE, 1975).

> It is proposed here that the determination of what mathematics is most worth learning is a task that will require careful and systematic study from the perspectives of several interest groups. (Helms & Graeber, 1975, p. 70)

> The challenge to describe basic skills and learning in school mathematics is an assignment full of pitfalls. In the past five years, hundred of mathematics educators, school systems, professional groups and the National Assessment have been busily composing taxonomies of fundamental objectives for mathematics instruction at various grade levels. With few exceptions, these efforts to establish a reasonable list of basic skills have been failures. There has been no general agreement among the competing groups. Moreover, the implementation of the various lists of curriculum guidelines threatens to produce fragmented mathematics programs that resemble occupational training more than they resemble education in mathematical methods and understandings likely to be of long range value. (Fey, 1975, p. 51, emphasis added)

These educators may have exaggerated the differences of opinion concerning composition of the basic skills, but it cannot be assumed without argument that one set of skills has general sanction.

Empirical Undimensionality

Empirical studies of the standardized tests seem to contradict the findings of our content analysis. High internal consistency has been reported for virtually all subtests, which would seem to indicate that the

test developers have been successful in constructing unidimensional tests. Developers of the SAT tests report, for example, that the mathematics subtests of concepts, computation, and applications given to beginning fifth graders (Intermediate Level 1, Form A, 1973) have internal consistency reliabilities of .87, .91, and .93, respectively.[5] Evidence of internal consistency has been taken as evidence that all items measure a single trait, and, as such, brings into question the utility of identifying subsets of items (see, for example, Goolsby, 1966).

There are at least two reasons, however, why conclusions based upon evidence of empirical unidimensionality may be misleading. The first stems from the definition of empirical unidimensionality; the second is a function of the ways in which unidimensionality is estimated.

The empirical definition of unidimensionality calls for a large first factor on the item intercorrelation matrix. Thus, empirical unidimensionality is a static concept specific to the time of test administration and to the population of respondents. For purposes of illustration, consider a population of respondents and a set of items that yield an item intercorrelation matrix with equal off-diagonal elements. Let the respondents be beginning fourth-grade students, and let half the items require division with remainder and half the items require multiplication of three-digit numbers. An intervention focused exclusively on multiplication of three-digit numbers might uniformly reduce the difficulty of half the items. In this case, the only effect the intervention would have on the item intercorrelation matrix would be to create a difficulty factor. Yet, despite empirical uni-

---

[5]These figures are reported in the SAT Manual, Norms Booklet, Form A, p. 15 (copyright 1973 by Harcourt Brace Jovanovich, Inc.).

dimensionality -- both prior to and after the intervention -- it is clear that there is a useful distinction between the two subsets of items. It is necessary, therefore, to ask whether or not a test is unidimensional relative to differences in instruction (i.e., Does all fourth-grade mathematics instruction affect all item difficulties equally?). Searching for differential effects across items is analogous to searching for aptitude-by-treatment interactions (ATIs) and might be called the "search for item-by-treatment interactions" (ITIs).

Most test data, however, are not obtained from people receiving uniform instruction. Different students have different educational experiences, and these experiences may have different effects across items. If a test consists of sets of items defined by concepts such that within each set the effect of an intervention is constant, and if the effects of interventions vary with less than perfect correlation across sets of items, the sets of items should be reflected in the pattern of item intercorrelations; the intervention effects contribute to both the covariance and variance of items within a set but not to the covariance of items in different sets.

Estimates of internal consistency based on data from norm groups of standardized tests seem to challenge the importance of ITIs. The apparent unidimensionality of standardized tests, however, might only be evidence of the existence of a strong single dimension, not of the absence of content factors. The Spearman-Brown prophecy formula implies that the more concepts included, the stronger the general factor, and that the fewer items per concept, the less clearly defined the second-order concept factors. Evidence of an internally consistent test should not be taken as an indication that searching for ITIs in evaluations using that test is useless.

Effects on Classroom Instruction

Critics of testing are concerned about the global effects of testing programs, and the analyses above suggest that some effect specific to the test selected may also exist. But effects of testing must also be considered in the broader investigation of the way instructional content is affected by factors outside the classroom.

Many believe that testing programs have some influence on the content teachers present. The prevailing opinion is that teachers are apt to "teach to the test," that is, to present content that closely follows the content of the test (see, for example, Cooley & Lohnes, 1976). Two groups consider this phenomenon undesirable: those who believe that the tests represent only a fraction of the content that is important, and those who believe that teachers should exercise their own judgment in determining instructional content. On the other hand, those who believe that the content of instruction should be uniform across classrooms (perhaps focusing on the basic skills) may see testing programs as a valuable tool for determining classroom instructional content. Each group bases its assessment of the tests, in part, on the assumption that teachers teach to the test, but neither can provide much empirical evidence to support that assumption.

It may be that the institution of a testing program leads teachers to spend more time on the material in the general area of test coverage, but not on the specific items covered by the test. Teachers, for example, might pay attention to the titles of the subtests but rely on their own preconceived ideas about the content in those areas to determine what content to cover. On the other hand, they might consider the testing program an unwarranted imposition and pay no attention to it

at all. In short, while the notion is plausible, it is far from certain that teachers teach to the specific content covered by tests used in their districts.

The likelihood that teachers teach to the test seems even smaller when one considers the other factors that might influence content choice. An alternative influence frequently cited in the literature is the textbook supplied to the teacher. Schutz (Note 5) has indicated that while teachers may initially claim that they select content by considering abstract goals of instruction, when pressed for detail, they often admit that they teach whatever material is covered in the textbook used by their students. Lists of objectives issued by the school district are another likely source of influence. The objectives are generally intended to influence, if not determine, content choice; hence, they act as a strong competitor to testing programs. It is not difficult to add to the list of possible influences. Teacher conceptions of subject matter, teacher assessment of student achievement, and student interest in subject matter all might affect a teacher's choice of content. When these alternative, partial determinants of content are considered, the assumption that teachers teach to the test seems less reasonable.

Advice to School Districts

Implicit in this discussion are two suggestions which might help to guide school districts in their selection of standardized tests.[6] (These suggestions apply to the question of whether <u>any</u> testing should be conducted only insofar as they may help administrators determine the

---

[6]The recommendations offered here are general; more specific recommendations must await the results of research currently underway at IRT. (Porter, Note 6; Floden, Note 7; Schmidt, Note 8, Freeman, Note 9.)

probable impacts of existing tests.)

First, the multitude of possible influences on content choice suggests that consistency is a minimal requirement if a district is to exercise control over instructional content in the classroom. If district administrators want specific content presented throughout the district, they would be well advised to ensure that the tools at their disposal are used toward that end. If textbook selection, test selection, and lists of objectives are all determined at the district level, then the administrator should make choices which will bring about consistent pressure on teachers to teach the desired content. To avoid losing ground by, say, choosing a textbook that covers different material from a carefully chosen test, administrators must make a thorough examination of tests, texts, and objectives. Furthermore, they should use whatever control they have over possible influences (such as parent pressure and principal pressure) to make sure that each emphasizes the same desired content.

Second, even if tests are ultimately found to have little specific influence on instructional content, it is still important that the content covered by the test match the content of most concern to those using the test results. The standardized tests do not all measure achievement in the same content areas, and the use of a test that assesses progress in an area of little concern to the district may be misleading.

Conclusion

The effects of test selection, both on the interpretation of results and on the content of instruction, have been over-simplified, denied, or ignored. Such treatment seems inappropriate when one considers that

an examination of four major tests has raised serious doubts about the assumption that the tests all measure the same achievement. If this widely-held assumption does not stand up to scrutiny, there is ample reason to question other assumptions about the effects of test selection.

Consideration of the effects of tests on content immediately calls to mind a broader area much in need of study -- the manner in which teachers choose instructional content. Research is needed to identify the mechanisms by which teachers respond to the multitude of pressures to choose instructional content. A first attempt at investigating this area is now underway at the Institute for Research on Teaching. In two parallel studies, the relative influences of six external factors are being examined. The factors are: testing program, set of objectives, textbooks, pressure from parents, pressure from teachers of higher grades, and pressure from the principal. In one study, teachers are asked to indicate how they think they would react to these pressures in a hypothetical situation. In the second study, assessments are being made, in a number of school districts, of the relationship between content covered and the external pressures at work (Porter, Note 6; Floden, Note 7; Schmidt, Note 8; Freeman, Note 9). Both studies deal only with fourth-grade mathematics. While these studies should provide some clues about the ways in which teachers determine content choices, the process is surely complex and will require prolonged study before it is well understood.

If content covered is important in terms of student achievement (and this is generally acknowledged), then the means by which teachers choose content, and the means by which administrators can influence that choice must be better understood. Indeed, increased understanding in these areas could be the key to the much sought improvement in student performance.

Selecting tests on the basis of the content they cover may provide immediate benefits and can serve as a starting point for the investigation of the relationships among testing, achi ement, and content coverage.

Reference Notes

1. Wiley, D.E. Policy-responsive evaluation. Paper presented at the Center for the Study of Evaluation Winter Invitational Conference on Measurement and Methodology, Los Angeles, January 1978.

2. Porter, A.C., Schmidt, W.H., Floden, R.E., & Freeman, D.J. Impact on what?: The importance of content covered. (Res. Ser. No. 2). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978. Paper originally presented at the annual meeting of the Evaluation Research Society, Washington, D.C., October, 1977.

3. Schmidt, W.H., Porter, A.C., Floden, R.E., & Freeman, D.J. Training manual for the classification of the content of fourth-grade mathematics (Res. Ser. No. 4). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978.

4. Linn, R.L. Evaluation of Title 1 via the RMC models: A critical review. Paper presented at the Center for the Study of Evaluation Winter Invitational Conference on Measurement and Methodology, Los Angeles, January 1978.

5. Schutz, R. Measurement in instruction. Paper presented at the Center for the Study of Evaluation Winter Invitational Conference on Measurement and Methodology, Los Angeles, January 1978.

6. Porter, A.C. Factors affecting teachers' decisions about content. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

7. Floden, R. A policy-capturing study of teacher decisions about content. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March, 1978.

8. Schmidt, W.H. Measuring the content of instruction (Res. Ser. No. 35). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978. Paper originally presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

9. Freeman, D.J. Conceptual issues in the content/strategy distinction (Res. Ser. No. 21). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1978. Paper originally presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

References

Cooley, W.W., & Lohnes, P.R. Evaluation research in education. New York: Irvington, 1976.

Fey, J.T. Remarks on basic skills and learning in mathematics. In The NIE Conference on Basic Mathematical Skills and Learning, Vol. 1: Contributed position papers. Washington, D.C.: NIE, 1975.

Goolsby, T.M. Differentiating between measures of different outcomes in the social studies. Journal of Educational Measurement, 1966, 3, 219-222.

Helms, D., & Graeber, A. Problems related to children's acquisition of basic skills and learning of mathematics and some suggested R&D options for NIE support. In The NIE Conference on Basic Mathematical Skills and Learning, Vol. 1: Contributed position papers. Washington, D.C.: NIE, 1975.

Linn, R.L., & Slinde, J.A. The determination of the significance of change between pre and posttesting periods. Review of Educational Research, 1977, 47, 121-150.

National Institute of Education. The NIE Conference on Basic Mathematical Skills and Learning, Vol. 1: Contributed position papers. Washington, D.C.: NIE, 1975.

Walker, D.F., & Schaffarzick, J. Comparing curricula. Review of Educational Research, 1974, 44, 83-112.